Slide 1 of 17

Title Slide: Data Considerations for Analysis

Slide 2 of 17

This module describes the analytic considerations that should be kept in mind when using HSLS:09 data. It describes the similarities and differences of HSLS:09 with past high school longitudinal studies. It provides information about state representative data and student psychosocial measures. This module also discusses the considerations researchers should take into account when using the data files, as well as the differences between public-use and restricted-use data files.

Slide 3 of 17

HSLS:09 is the fifth in a series of secondary school longitudinal studies sponsored by NCES. The design is similar to past studies such as High School and Beyond, or HS&B, National Education Longitudinal Study of 1988, or NELS:88, and Education Longitudinal Study of 2002, or ELS. These studies describe the educational experiences of students from three decades: the 1980s, 1990s, and 2000s. They also provide bases for further understanding the correlates of educational success in the United States. The next two slides highlight the similarities and differences of these studies with HSLS:09.

Slide 4 of 17

HSLS:09 has major points of continuity with all or several of the past NCES secondary school longitudinal studies. All five longitudinal studies share a commitment to collecting high school (grades 9-12) transcripts across a nationally representative school sample with an oversample of private schools with student numbers that are sufficient for subgroup reporting by major race/ethnicity categories, including Asians. Across all five studies there is a shared commitment to following the cohort beyond high school as well as to identify and follow high school dropouts. The studies all provide contextual samples of parents and teachers, and include a school administrator survey. The HSLS:09 and NELS and ELS also share an ability-adaptive assessment battery. Finally, all of the NCES secondary school longitudinal surveys produce a general purpose dataset that supports a broad range of descriptive and interpretive reporting.

Slide 5 of 17

HSLS:09 has important differences with the prior studies. Shown here are the distinctive and innovative features of HSLS:09. These include: (1) the use of a computer-administered assessment and student questionnaire in a school setting; (2) an assessment that focuses on algebraic reasoning; (3) use of computerized (web/computer-assisted telephone interview) parent, teacher, administrator, and counselor questionnaires; (4) inclusion of a counselor survey to document school course and program assignment policies and procedures; (5) a starting point in the fall of ninth grade, which marks the traditional beginning of high school; (6) enhanced emphasis on the dynamics of educational and occupational decision-making; (7)

enhanced emphasis on science, technology, engineering, and mathematics (or STEM) trajectories; and (8) follow-up in spring of students 11th grade year which includes a follow-up mathematics assessment.

Slide 6 of 17

Additionally, the HSLS:09 provides researchers with the ability to explore general trends in youth transition, not grade-based specific comparisons with prior spring cohorts of eighth-graders, sophomores, and seniors as in other past studies. HSLS:09 also provides a linkage to selected state administrative data systems and augmentation of selected state public school samples to render them state-representative.

Slide 7 of 17

The nationally representative HSLS:09 data set can also be representative of public school ninth graders in each of ten states. These states are California, Florida, Georgia, Michigan, North Carolina, Ohio, Pennsylvania, Tennessee, Texas, and Washington.

Slide 8 of 17

As one example, in North Carolina, approximately 88 percent of public school 9th-graders were proficient in algebraic expressions, 64 percent were proficient in multiplicative and proportional thinking, 46 percent had mastered algebraic equivalents, 22 percent had mastered systems of equations, and 11 percent were proficient in linear functions.

Slide 9 of 17

The HSLS:09 restricted-use data file provides a variable, X1STATESAMPLE, that researchers can use to limit the main data set to be reflective of each of the ten states. As stated, even though the full HSLS:09 national sample includes Catholic and other private schools and their students, the state samples are representative of **public school** ninth graders. The values for this variable are shown here.

Please refer to Common Module 5, ‘Acquiring Micro-level NCES Data’ for more information on obtaining restricted-use data.

Slide 10 of 17

The ECB allows users to extract SAS, SPSS, and Stata programs. The extract program functionality allows users to create an ECB taglist of variables and extract those variables to a program that can be subsequently run. The X1STATESAMPL variable should be moved to the taglist so that it is available in the program. Users can modify the program to subset the data to a particular sample. The example shown here is the sample code for each type of program using the North Carolina state public school sample as an example. Please note that X1STATE should not be confused with X1STATESAMPL when determining records in the state sample. X1STATE applies to

all records and indicates the state that the school resides, where X1STATESAMPL applies to records specific to the public school state sample.

Slide 11 of 17

Researchers interested in making state level estimates for the 10 states oversampled within HSLS:09 use the same analytic weights as they would for making national estimates. When the data are filtered or subset, the sum of the weights associated with that subset reflect the population of interest. In this case, the population of interest is the number of public school 9th graders within the state.

Slide 12 of 17

Within the 944 schools that participated in HSLS:09, questionnaire responses and mathematics assessment scores were obtained from 21,444 9th-grade students. This table provides information on the approximate number of public schools that responded, by state and the unweighted and weighted response rate, by state. Column A lists the ten states for which state-representative data are available from the public school sample as well as a total of all public-schools and the other states without state-representative data. Column B provides the number of responding schools rounded to the nearest ten, as these are restricted-use data. Columns C and D provide the unweighted and weighted percent of responding schools, respectively.

Slide 13 of 17

This table provides information at the student level on the approximate number of public school ninth-graders that responded, by state and the unweighted and weighted response rate, by state.

Column A lists the ten states for which state-representative data are available from the public school sample as well as a total of all public-schools and the other states without state-representative data. Column E provides the number of responding students with questionnaire or assessment data within responding schools, rounded to the nearest ten, as these are restricted-use data. Columns F and G provide the unweighted and weighted percent of responding students with questionnaire/assessment data within responding schools, respectively.

Slide 14 of 17

The base year of HSLS data provide several measures that assess students' attitudes and experiences. For example, the survey includes questions on student self-efficacy in math and science, their attitudes about school, math, and science, and asks whom students talk to about education and career plans. These data are available for both the individual survey items, as well as in the form of composite scores, such as mathematics self-efficacy.

As mentioned, the HSLS:09; NELS:88; and ELS:2002 have similarities in terms of study design. Despite differences in emphasis, all of the studies draw content from the same

or similar theoretical constructs (for example, achievement growth, school effectiveness, social capital, social attainment, human capital). Therefore, researchers can examine these constructs across the different data sets.

Slide 15 of 17

There are some considerations that researchers should take into account with using HSLS:09 data. One such consideration is that the teacher data are not representative of all teachers. It is important for researchers to keep in mind that the teacher data supply contextual information for students, and that the students are the unit of analysis. Thus, the resulting teacher sample is not representative of teachers in any given school, state or the US as a whole. The design of this data component does not provide a standalone analysis sample of teachers, but instead permits specific teacher characteristics and practices to be related directly to the learning context and educational outcomes of sampled students.

Researchers should also be aware that for the first follow-up, parent interview data were collected from a subsample of parents. The HSLS:09 analytic weighting scheme accounts for this subsampling. Researchers are encouraged to read Chapter 8 of the First Follow-up Data File Documentation for more guidance on the weights that support analyses with parent data.

The last consideration that researchers should take into account is that HSLS:09 does not offer specific information on special education services or data from special education teachers. Researchers who are interested in topics related to special education should consult other studies available, such as National Longitudinal Transition Study-2 (NLTS2), that are specific to special education services.

Slide 16 of 17

NCES provides two types of data files: public-use data and restricted-use data files. While the public-use data files are readily available on the HSLS website, restricted-use data files contain more detailed information than the public-use files. For instance, some student and school data are suppressed from that public-use file, which can be found in the restricted-use file. Additionally, you cannot do school- and student-level data analysis using public-use HSLS:09 data files. Lastly, as the PSU and strata identifiers are suppressed in the public-use data file, only replication techniques can be used for complex sample variance estimation.

In addition to this information being included in the restricted-use data files, they also include state representative data for ten states, school characteristics, and other more detailed information than what is available in the public-use dataset. Since restricted-use data files may be considered sensitive or unique, researchers who are interested in accessing restricted-use data must apply for, and be granted, a restricted-use data license from NCES. For more information regarding access to restricted-use data files please refer to Common Module 5: Acquiring Micro-level NCES Data, which can be accessed by clicking the underlined screen text ‘restricted-use’.

Slide 17 of 17

This module has provided you with analytic considerations that should be kept in mind when using this data. Links to resources are provided here for your reference. You may now proceed to the next module in the series, or exit the module.